CCRIFX PROJECT DESCRIPTION – DRAFT (microarrays)

Project Number: CCRIFX-
Project Title:
Requestor:
Requestor Lab PI:
Address:
Additional Investigators to include:
Bioinformatics contact:
Completed Request(s):
Current Active Request(s):

Rationale/Significance:

Scientific questions or goals (choose all that apply)

- **Transcriptional profiling**: aimed to estimate absolute genes expression values in a cell type/tissue
- **Identification of gene expression signatures:** aimed to identify expression patterns uniquely characteristic of a medical or other condition such as tumor stage, subtype, aggressiveness, metastatic potential, etc.
- **Class comparison**: aimed to determine whether gene expression profiles differ among samples from different classes and to identify which genes are differentially expressed between classes.
- **Testing for interaction between classes:** aimed to identify synergistic effects between two factors (e.g. two drugs).
- **Class prediction**: aimed to develop a statistical model that can predict which class a new specimen belongs to based on its expression profile.
- **Class discovery**: aimed to classify samples and/or to identify novel subtypes of specimens within a population. E.g. which gene expression signatures are associated with prolonged survival?
- **Enrichment analysis**: aimed to identify differentially regulated pathways
- **Gene network reconstruction:** aimed to do reverse engineering of gene regulatory networks

Approach

- *E.g. Microarray gene expression profiling of murine fibroblasts (KO and WT) treated with …*

Impact

- *If successful, this study will identify the role of XXX in immune response…*

Other priority considerations

- Event(s) in the near future that makes this request time-sensitive:
  - Site visits
  - Manuscripts/Proposals
  - Posters/Presentations
- The availability of data to be analyzed:

Template revised July 18, 2013

    - Not available yet
    - Available now
    - Public sources
- Role of the project:
    - Publishable experiment
    - Exploratory
    - Confirmatory

Known risks and limitations

Gene expression microarray profiling projects can have dramatically different goals. To maximize the chance of meeting the goals, it is critical to use appropriate experimental design.

- The following factors can help reduce bias and variability in microarray experiments:
    - Ideally sample collection and NA extraction should be performed the same day by the same individual using the same protocol and reagents.
    - If that is not possible, samples should be frozen and processed together at a later date at the same microarray facility. For large studies, randomization and blocking should be used when applicable.

- At least three replicates per condition are recommended to allow for statistical analysis. More is better, because it will improve the sensitivity of the analysis.

- More replicates are recommended for experiments that involve
    - High biological variability (e.g. human or mouse samples vs. cell lines)
    - Non-target tissue contamination (e.g. mouse embryonic tissues)
    - Subtle treatment effect (e.g. weak transcriptional signal)
    - Multiple treatments
    - The goals include understanding the mechanism of action or network analysis

- Note: For more guidance, see the checklist for Nature journals:
    - www.nature.com/authors/policies/checklist.pdf

Original description

[Text describing the analysis goes here]

Experimental design and metadata

- **Samples**
    - Species and strains:
        - Human (race if known)
        - Mouse (e.g. C57BL/6, FVB/NJ, 129P2/OlaHsd)
        - Other (please specify)
    - Type: tissue, cell lines, cells, etc.
    - Extracted material: non/polyadenylated long RNA, small RNA, miRNA, etc.
    - Total number of
        - Samples -
        - Groups -

        - Replicates (biological and technical) -
    - Diversity of the population, intra-individual variability of the sample, clonal heterogeneity, similarity to the reference, non-target tissue contamination, etc. – if known

- **Protocols**
    - GEO submission fields (Excel file)
    - Treatment and enrichment protocols used in sample prep:
    - Microarray platforms used:
        - Affymetrix
        - cDNA or spotted arrays
        - Other one-color arrays: Agilent, Applied Biosystems, Eppendorf, GE Healthcare, Illumina, Nimblegen
        - Other two-color arrays: Agilent, NCI_Operon
        - Pathway-Specific PCR arrays
- **Data**
    - Location and format: cel, chp and grd files
    - SAIC-F CSAS#
    - Preferred genome references: version/source (e.g. hg19, mm9, mm10)
    - For public data mining requests:
        - Diseases, cancer subtypes, databases: ENCODE, GEO, etc.
        - Types of data: RNA-Seq, ChIP-Seq, microarray, etc.

Analysis Details

- **Analysis desired from CCRIFX**
    - Preferred analytical applications – standard or custom
    - Major steps in the workflow or procedure that will be followed
        - Normalization and expression value estimation
        - QC: PCA and Correlation plots, R coefficient across replicates
        - QC tools: Affymetrix Power Tools; R/Bioconductor – Simpleaffy or arrayQualityMetrics (platform agnostic)
    - Differentially expressed genes (DEG) analysis
        - Software tools
            - BRB Arraytools
            - GeneSpring
            - Partek
            - LIMMA package
            - Affymetrix GCOS
        - Comparisons and controls
            - DE analysis
            - Testing for interaction between factors
    - Genome features of interest: genes, transcripts, exons

- **Results expected from CCRIFX**
    - Files: gene lists, pathway lists, etc.
    - Figures: heatmaps, venn diagrams
    - List of methods used with brief descriptions

    - Education and training
    - Submission to public repositories: GEO, GenBank

Deferred tasks

- Related analyses on the same data that are too big to fit into the current request
- Related analyses on that require additional data (e.g. data integration)
- Requests for continuous support

Publications

- References specific approaches, analyses or protocols that the investigator wants to replicate
- Publications that provide scientific context for the analysis